**Assignment #1 - Group 3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Student ID** | **Section Contributed** | **Section Edited** | **Other Contributions** |
| Nash, Emma | 301452134 | GitHub syncing, Code for sub corpora #3 | Code | Group organization |
| Porime, Catalina | 301440736 | Corpus collection,  GitHub syncing, Code for sub corpora #2 | Code | Code assistance |
| Changfoot, Kinzlee | 301425282 | Corpus collection,  GitHub syncing, Code for sub corpora #1 | Code | Code assistance |
| Jagani, Zayn-Ali | 301469969 | Joining code, introduction write up | Write up | Coordination |
| Anderson, Kaleb | 301562090 | Joining code, response to Q7 | Write up | Coordination |

**Link to Repository on GitHub:**

<https://github.com/cataporime/Assignment1_Group3>

Here you can find our txt files that we analyzed, our Jupyter notebook files, as well as the PDF of our code and output.

**Introduction:**

Jeffrey Dahmer, a convicted serial killer, has been a subject of widespread interest in various media forms, including news, TV shows, and podcasts. His notoriety reached global audiences, and his life was even portrayed in a Netflix documentary. The media closely followed his trial due to the danger he posed to society. The following assignment explores the portrayal of Dahmer’s life from a statistical perspective, analyzing the word count, lexical diversity, and other textual features of different media genres. We chose Dahmer as our focus because of the extensive coverage his story has received, from his trial to the continued study of his life, even beyond his death. His case remains a topic of interest in fields ranging from psychology to criminology.

**Sub Corpora:**

1. Sub Corpora #1: Academic article on Jeffery Dahmer[Academic Article on Jeffery Dahmer](https://epublications.regis.edu/cgi/viewcontent.cgi?article=1240&context=theses)
2. Sub Corpora #2: YouTube documentary on Jeffery Dahmer <https://www.youtube.com/watch?v=Y1EWmrzD2Mk>
3. Sub Copera #3: Reddit thread discussion of Jeffery Dahmer <https://www.reddit.com/r/AskALiberal/comments/xr64cy/what_do_you_think_of_jeffery_dahmer/>

**Question 1-6:**

On separate PDF found on Canvas or in our GitHub repository.

**Question 7:**

The most frequent words, longest words, and longest sentences across the three sub corpora reveal distinct characteristics of each genre. Examining the first corpora, the academic article, the longest words reveal a more technical and formal approach with terms like 'neglectful', 'socioeconomic', "psychopathy" and "antisocial”, aiming to analyze Dahmer’s psychological profile. This is also supported by its high occurrence of specialized terminology and a low lexical diversity, indicating a repetitive use of field-specific language, along with the longest sentence which is an example of a highly technical description of Dahmer’s acts. Reviewing the second corpora, the YouTube documentary, one of the most frequent words is “Dahmer” which reflects the style of a narrative retelling of events, the high lexical diversity can be explained by having multiple speakers in the documentary and the longest sentence shows the storytelling style of this documentary. Now the third corpora, the Reddit thread, the most frequent words being “Share”, “Downvote”, “2y” and “ago” showcase the formatting of Reddit, having a share and downvote button, along with reddit displaying when posts were made, shows this thread was most active 2 year ago. There is also a moderately high lexical diversity reflecting the diverse viewpoints and spontaneous exchanges typical of online discussions. The longest sentence being 165 words long and having 5 commas reveals the lack of structure needed in Reddit compared to both the movie and the